

Developed): -0.184  
Berkembang (Developing): 0.001

$$r = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sqrt{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \sqrt{\sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2}}$$

$r$  = Pearson correlation coefficient

$n$  = Number of observations

$X_i$  = Value of the  $i$  - th independent variable (Year)

$Y_i$  = Value of the  $i$  - th dependent variable (Inflation)

$\sum X_i, \sum Y_i$  = Sum of the values of each variable

$\sum X_i^2, \sum Y_i^2$  = Sum of squared values

$\sum X_i Y_i$  = Sum of the products of  $X_i$  and  $Y_i$

# EXPLORATORY DATA ANALYSIS: COMPARATIVE TRENDS OF INFLATION IN DEVELOPING VS DEVELOPED COUNTRIES

Verlene Angelica Kiarra Blue  
Data Science & Analyst Enthusiast

Tools: Python, Pandas, Matplotlib, Seaborn

Dataset: [datahub.io/core/inflation](https://datahub.io/core/inflation) (World Bank)

Fokus: Comparative trend, correlation, and inflation stability analysis

# Research & EDA Objective

## Tujuan:

- **Mengeksplor data inflasi** global selama 10 tahun (2013–2023)
- **Membandingkan tren inflasi** antara negara berkembang dan negara maju.
- Menganalisis **arah hubungan inflasi terhadap waktu** menggunakan **korelasi Pearson**.

## Why?

- Inflasi merupakan **indikator utama kestabilan makroekonomi global**.
- Tren inflasi pasca-pandemi menunjukkan **perbedaan signifikan antar kelompok negara**.
- **Pemahaman berbasis data** dapat mendukung kebijakan ekonomi yang lebih adaptif.

# Dataset Overview

- Sumber: DataHub.io / World Bank
- Periode data: 1960–2023 (**analisis fokus: 2013–2023**)

## Variabel utama:

- Country
- Country Code
- Year
- Inflation (%)

## Pembagian kategori:

- Berdasarkan klasifikasi **World Bank** → Developed vs Developing

	Country	Country Code	Year	Inflation
0	Aruba	ABW	1985	403.225.805.628.628
1	Aruba	ABW	1986	107.396.640.826.829
2	Aruba	ABW	1987	364.304.545.817.706
3	Aruba	ABW	1988	312.186.849.610.723
4	Aruba	ABW	1989	399.162.804.604.575

Jumlah Data



**11.040 rows**

Klasifikasi  
Kategori



**WorldBank**

# Data Cleaning & Processing

## Filter Tahun 2013–2023 agar data relevan

```
df_10yr = df_recent[df_recent["Year"] >= df_recent["Year"].max() - 9]
```

### Filtering Process

```
#Daftar negara sudah sesuai dengan World Bank
developed_countries = [
    "United States", "Canada", "United Kingdom", "Germany", "France",
    "Italy", "Japan", "Australia", "South Korea", "Netherlands",
    "Sweden", "Switzerland", "Norway", "Denmark", "New Zealand"
]

developing_countries = [
    "Indonesia", "India", "Brazil", "Vietnam", "Philippines",
    "Nigeria", "Pakistan", "Bangladesh", "Egypt", "Kenya",
    "Peru", "Thailand", "Mexico", "South Africa", "Argentina"
]
```

```
df_dev = df[df['Group'] == 'Developed']
df_dev2 = df[df['Group'] == 'Developing']
```

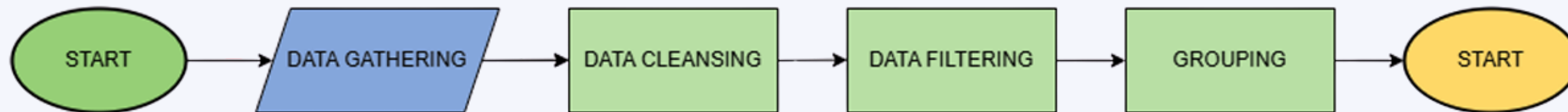
### Cleaning Process

```
# Hapus NaN pada kolom Inflasi
df = df.dropna(subset=["Inflation"])

# Hapus karakter yang bukan angka (misalnya koma atau titik ribuan)
df["Inflation"] = df["Inflation"].str.replace(",", "").str.replace(".", "")

# Konversi ke float
df["Inflation"] = pd.to_numeric(df["Inflation"], errors="coerce")

# Hapus baris yang masih tidak valid setelah konversi
df = df.dropna(subset=["Inflation"])
```





# Data Understanding

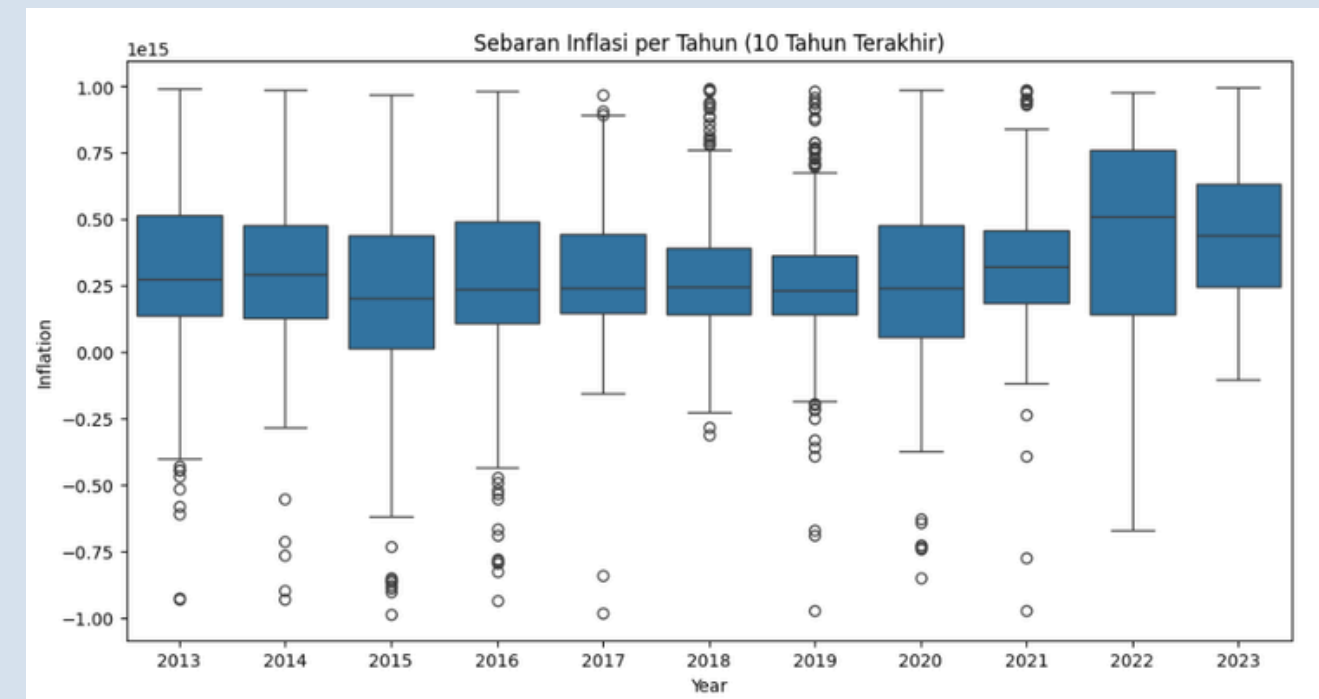
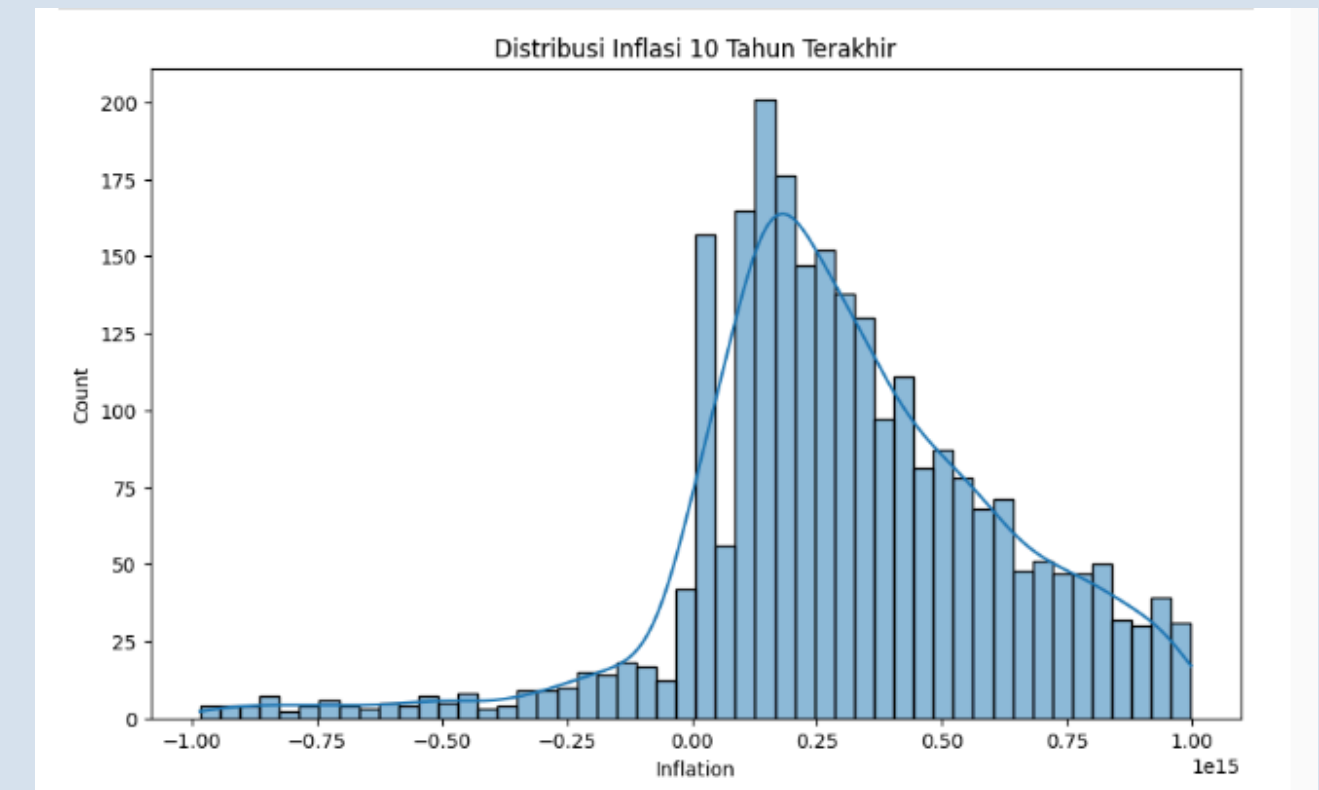
## Histogram Insight

- Mayoritas negara memiliki inflasi di bawah 10%.
- Distribusi condong ke kanan (right-skewed) → beberapa negara mengalami inflasi ekstrem tinggi.
- Pola ini menunjukkan bahwa sebagian besar negara relatif stabil, tapi masih ada outlier signifikan.

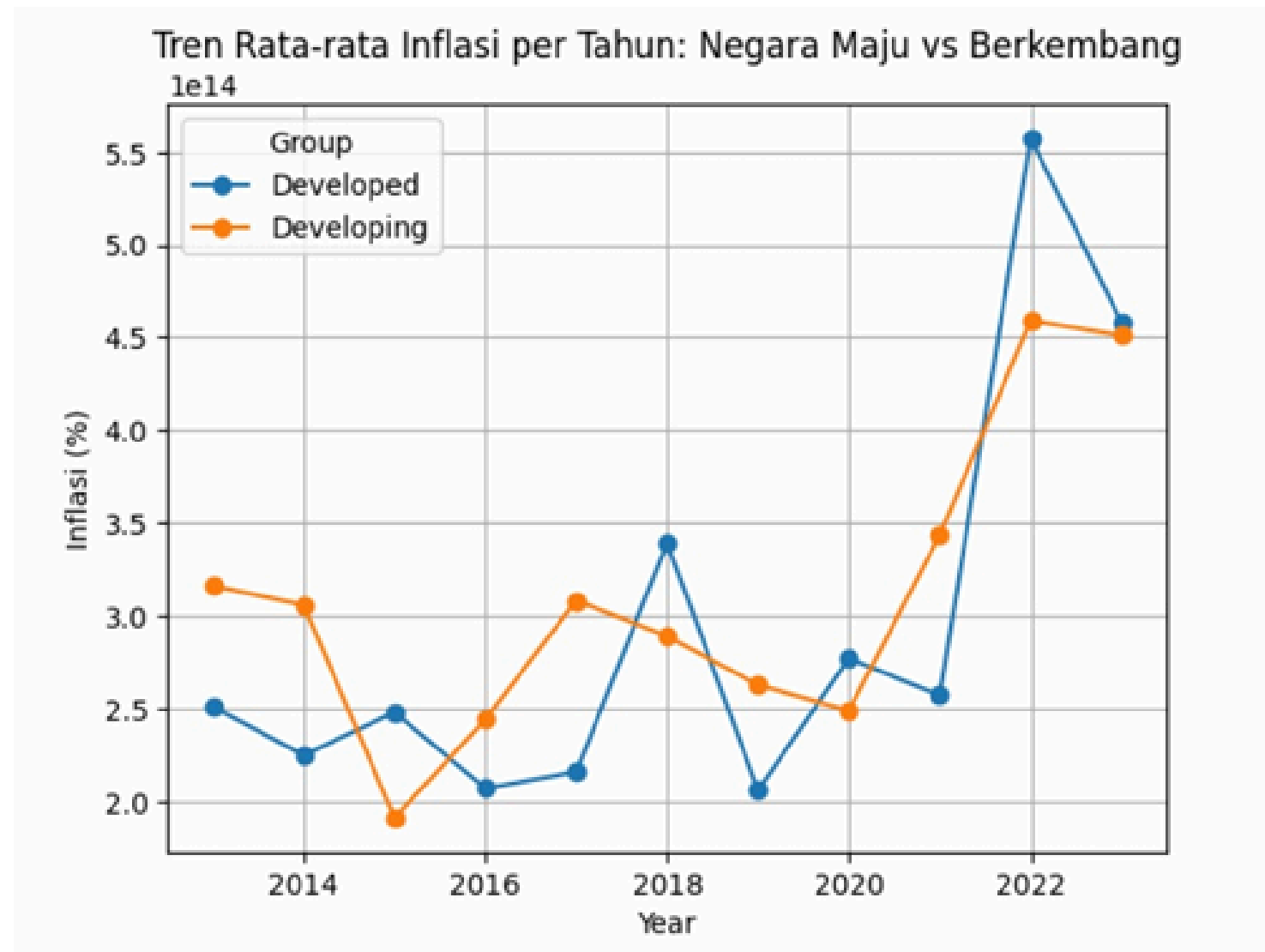
## Boxplot Insight

- Rentang nilai inflasi cukup lebar, menunjukkan variasi tinggi antar negara.
- Terdapat beberapa outlier ekstrem yang menandakan adanya negara dengan kondisi inflasi tidak terkendali.

Data menunjukkan ketimpangan signifikan antara negara maju dan berkembang >> dasar yang kuat untuk analisis tren dan korelasi berikutnya.



# Trend Analysis



Negara berkembang cenderung memiliki inflasi lebih tinggi dan fluktuatif dibanding negara maju yang lebih stabil.

# Heatmap Analysis

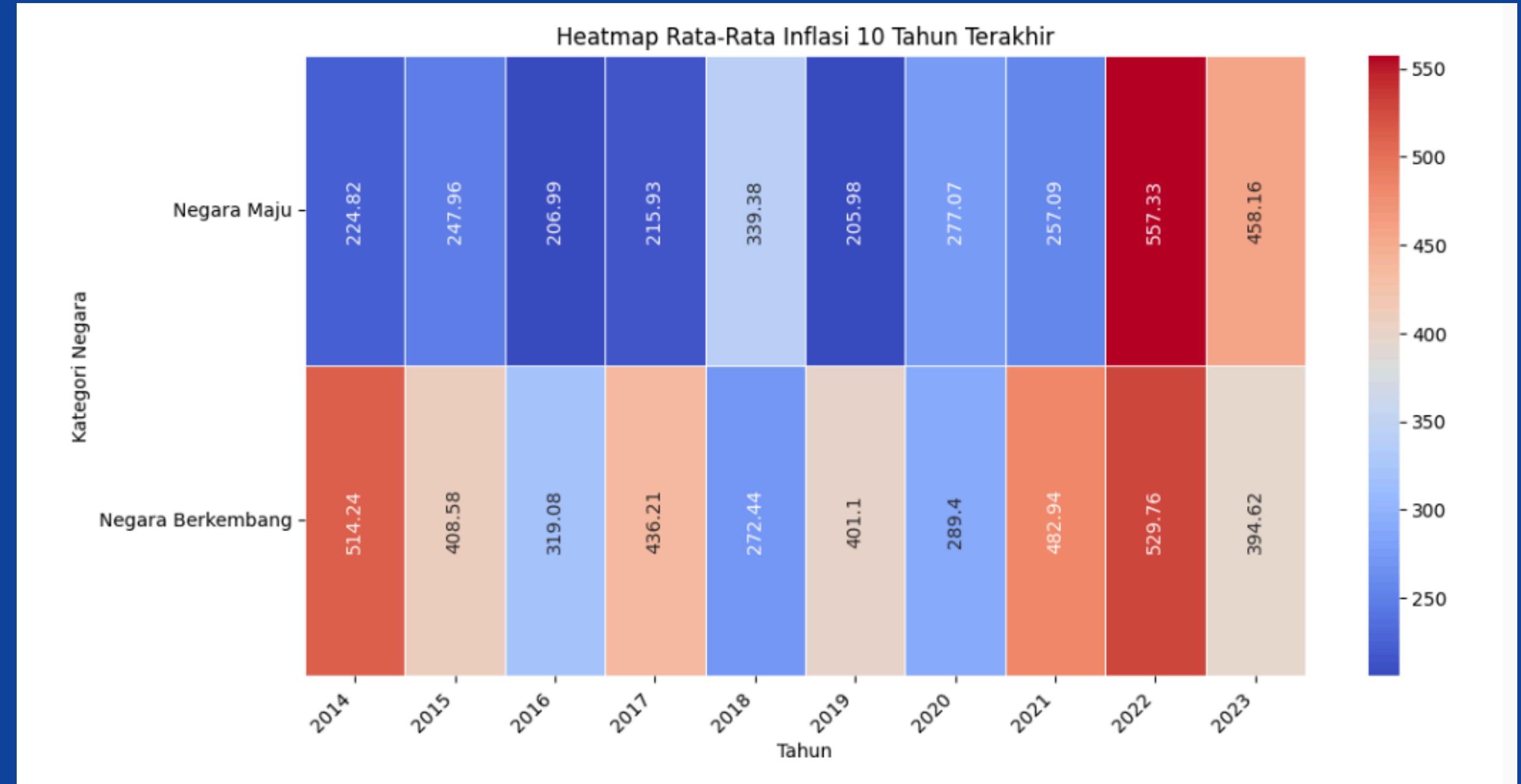
X-axis: Tahun

Y-axis: Kategori negara

Warna: Intensitas inflasi

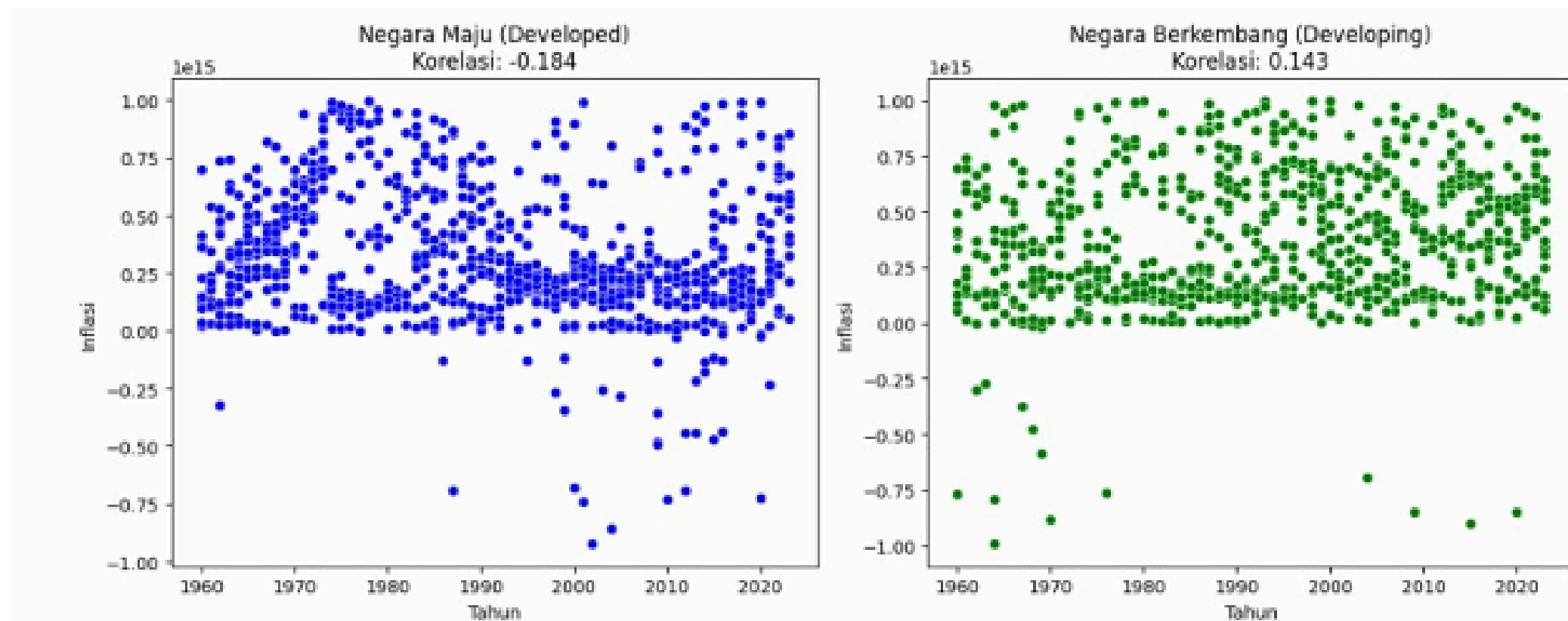
## Insight:

Tahun 2022 menunjukkan lonjakan inflasi global yang ekstrem, terutama pasca-pandemi.



# Correlation Analysis (Pearson)

$$r = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sqrt{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \sqrt{\sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2}}$$



```
from scipy.stats import pearsonr
```

```
r_dev, _ = pearsonr(df_dev['Year'], df_dev['Inflation'])  
r_dev2, _ = pearsonr(df_dev2['Year'], df_dev2['Inflation'])
```

```
print(f"Korelasi negara maju (Developed): {r_dev:.3f}")  
print(f"Korelasi negara berkembang (Developing): {r_dev2:.3f}")
```

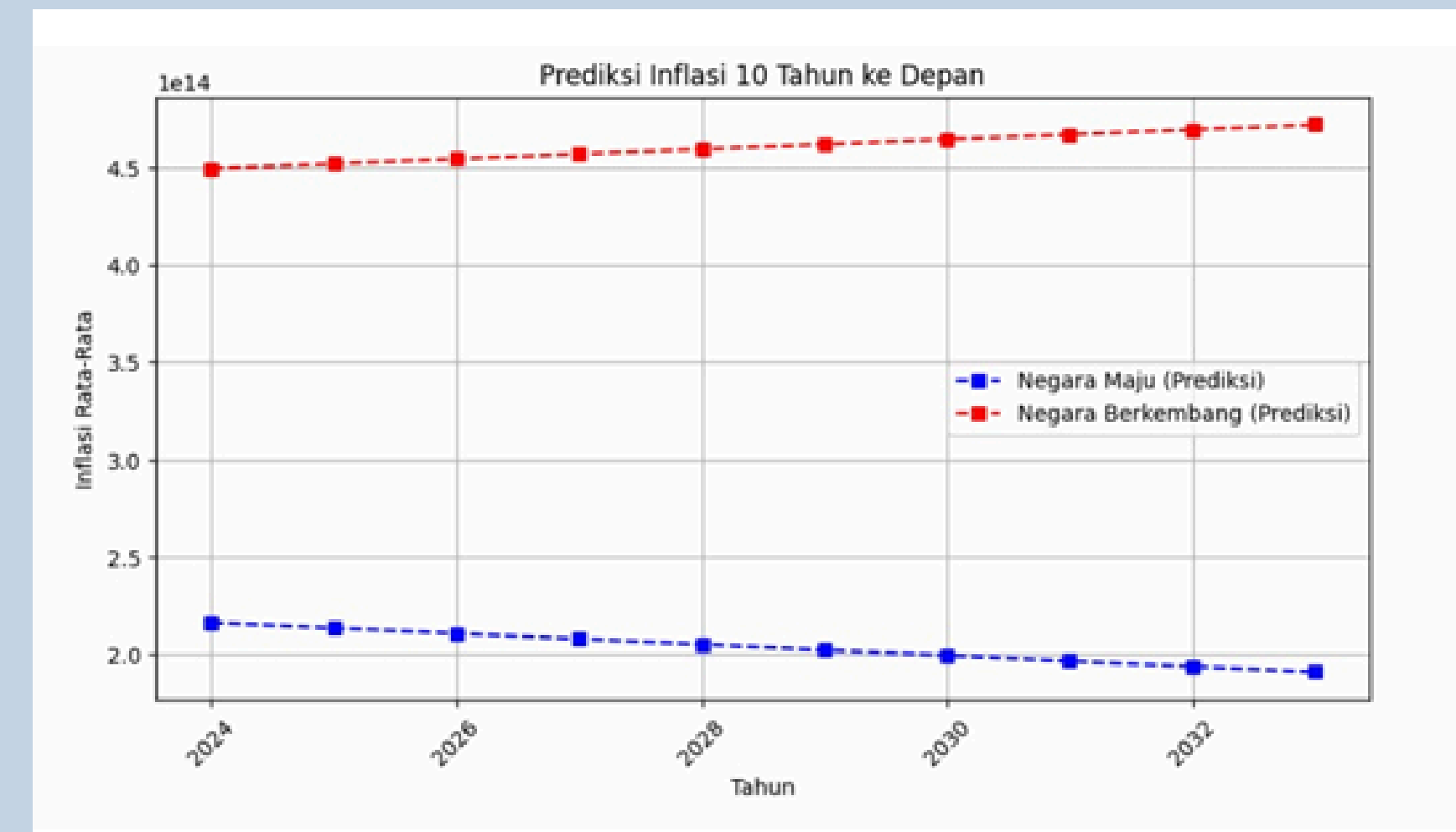
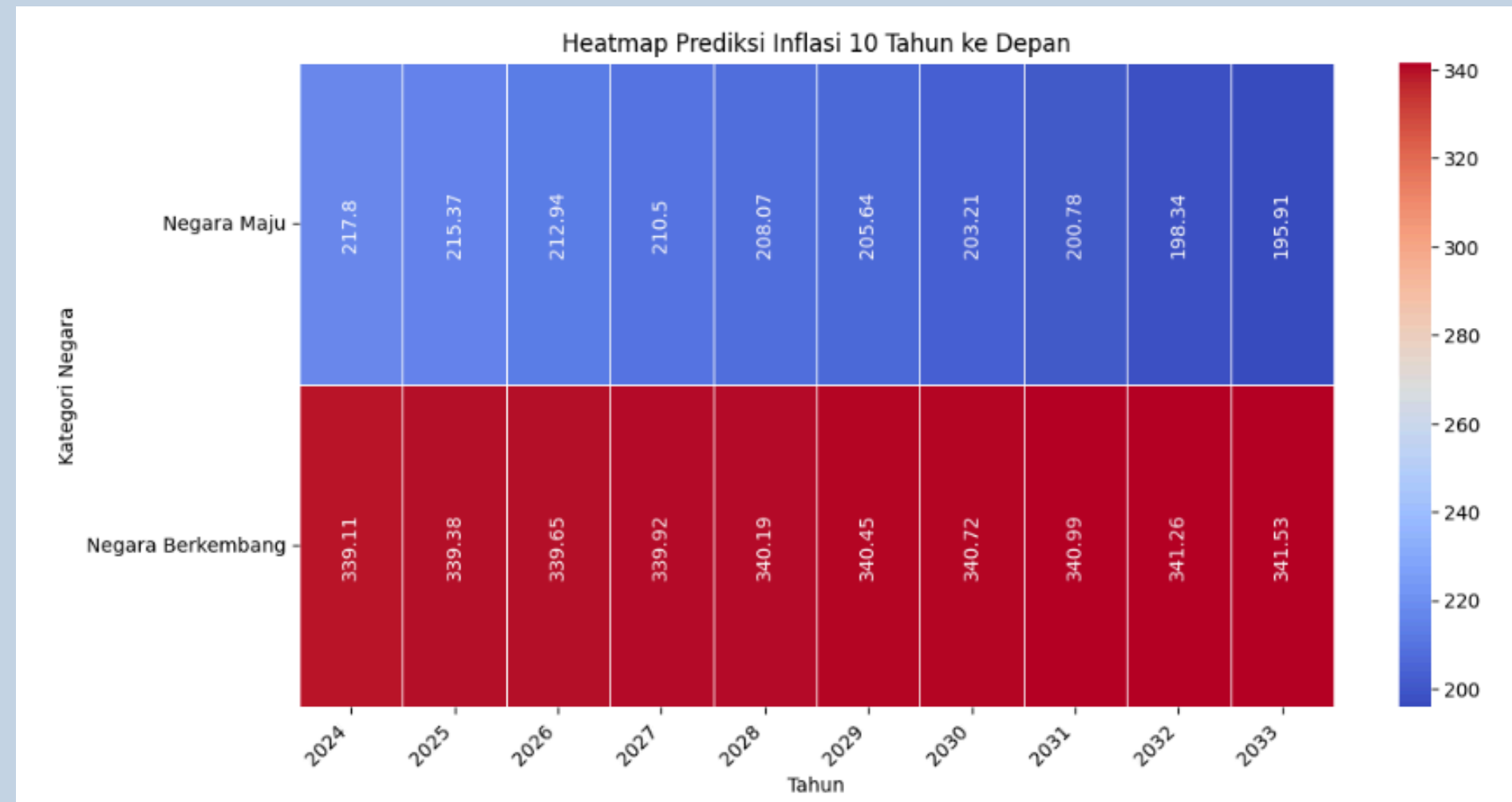
```
Korelasi negara maju (Developed): -0.184  
Korelasi negara berkembang (Developing): 0.001
```

Negara maju menunjukkan korelasi negatif, menandakan kemampuan kontrol harga lebih baik. Negara berkembang masih menghadapi tekanan inflasi jangka panjang.

Negara **Maju** : **r = -0.184** (negatif → inflasi makin stabil)  
Negara **Berkembang** : **r = 0.143** (positif → inflasi meningkat)



# Predictive Insight



Negara **berkembang** diproyeksikan **mengalami kenaikan inflasi**,  
sedangkan negara **maju stabil** atau **menurun perlahan**.

# Conclusion & Insight

- **Integrasi Big Data** dan **analisis visual** dapat mengungkap **pola makroekonomi secara jelas**.
- Negara berkembang butuh **kebijakan fiskal/moneter lebih adaptif**.
- Insight ini bisa **dikembangkan ke model multivariat** (GDP, exchange rate, dsb).

[Github Project Link](#)



[LinkedIn](#)



[verlenekiarra@gmail.com](mailto:verlenekiarra@gmail.com)



# Thank You!

Always Keep Learning,  
Stay Insightful with Me!

